#	LS

HKUST – Department of Computer Science and Engineering

COMP170: Discrete Math Tools for CS – FALL 2007

Final Examination

Date: Saturday Dec 15, 2007 Time: 4:30–7:00pm

Name: ______________________	Student ID: ______________
Email: ______________________	Lecture and Tutorial: ________

Instructions

- This is a closed book examination. It consists of 29 pages and 11 questions.
- Please write your name, student ID, email, lecture and tutorial sections on this page.
- For each subsequent page, please write your student ID at the top of the page in the space provided.
- Please sign the honor code statement on page 2.
- Answer all the questions within the space provided on the examination paper. You may use the back of the pages for your rough work. The last three pages are scrap paper and may also be used for rough work. Each question is on a separate page (and sometimes has an extra page for you to do work on). This is for clarity and is not meant to imply that each question requires a full page answer. Many can be answered using only a few lines.
- Only use notation given in class. Do not use notation that you have learnt outside of this class that is nonstandard.
- Calculators may be used for the examination.

Questions	1	2	3	4	5	6	7	8	9	10	11	Total
Points	5	6	10	7	11	10	10	9	11	11	10	100
Score												

Student ID: ______________

As part of HKUST's introduction of an honor code, the HKUST Senate has recommended that all students be asked to sign a brief declaration printed on examination answer books that their answers are their own work, and that they are aware of the regulations relating to academic integrity. Following this, please read and sign the declaration below.

> I declare that the answers submitted for this examination are my own work.
>
> I understand that sanctions will be imposed, if I am found to have violated the University regulations governing academic integrity.
>
> Student's Name: ______________________
>
> Student's Signature: ______________________

Student ID: ______________

Definitions and Formulas: This page contains some definitions used in this exam and a list of formulas (theorems) that you may use in the exam (without having to provide a proof). Note that you might not need all of these formulas on this exam.

Definitions:

1. $N = \{0, 1, 2, 3, \ldots\}$, the set of non-negative integers.
2. $Z^+ = \{1, 2, 3, \ldots\}$, the set of positive integers.
3. R is the set of *real numbers*.
4. R^+ is the set of positive *real numbers*.

Formulas:

1. $\binom{n}{i} = \frac{n!}{i!\,(n-i)!}$
2. If $0 < i < n$ then $\binom{n}{i} = \binom{n-1}{i-1} + \binom{n-1}{i}$.
3. $\neg(p \wedge q)$ is equivalent to $\neg p \vee \neg q$.
4. $\neg(p \vee q)$ is equivalent to $\neg p \wedge \neg q$.
5. $\sum_{i=1}^{n-1} i = n(n-1)/2$
6. $\sum_{i=1}^{n-1} i^2 = \frac{2n^3 - 3n^2 + n}{6}$
7. If $r \neq 1$ then $\sum_{i=0}^{n-1} r^i = \frac{1-r^n}{1-r}$.
8. If $r \neq 1$ then $\sum_{i=0}^{n} ir^i = \frac{nr^{n+2} - (n+1)r^{n+1} + r}{(1-r)^2}$.
9. The inclusion-exclusion theorem:
$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{\substack{i_1, i_2, \ldots, i_k: \\ 1 \le i_1 < i_2 < \ldots < i_k \le n}} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_k})$$
10. If X is a random variable, then $E(X)$ denotes the *Expectation of* X and $V(X) = E\left((X - E(X))^2\right)$ denotes the *Variance of* X.
11. $f(n) = O(g(n))$ if there exist some $N > 0$ and positive constant c such that $\forall n > N$, $f(n) \le c \cdot g(n)$.
12. $f(n) = \Theta(g(n))$ if $f(n) = O(g(n))$ and $g(n) = O(f(n))$.

Student ID: ______________

Problem 1: [5 pts]

For each of the multiple-choice problems below, **circle** the correct answer. For this question, no work needs to be shown. In what follows S_n is a set containing n items.

(a) Assume $m > n$. The number of *one-to-one* functions from S_n to S_m is

(i) 0 (ii) n^m (iii) m^n (iv) $n^{\underline{m}}$ (v) $m^{\underline{n}}$

(b) Assume $m > n$. The number of *onto* functions from S_n to S_m is

(i) 0 (ii) n^m (iii) m^n (iv) $n^{\underline{m}}$ (v) $m^{\underline{n}}$

(c) How many elements in Z_{91} have a multiplicative inverse?

(i) 18 (ii) 70 (iii) 71 (iv) 72 (v) 62

(d) Suppose the logical statement $\neg(p \Rightarrow q)$ is `TRUE` .
Which of the following statements is then always also `TRUE` ?

(i) $q \vee \neg p$ (ii) $\neg q \wedge \neg p$ (iii) $\neg q \vee p$ (iv) $q \wedge p$
(v) None of the above

(e) Consider the following three statements:
A) $\forall x \in R^+, \quad x^3 \geq 4 - 3x$
B) $\forall x \in Z^+, \quad x^3 \geq 4 - 3x$
C) $\forall x \in N \left(\forall y \in N, \quad ((x \geq y) \wedge (y \geq z)) \Rightarrow ((x - z) > (x - y)) \right)$
Which of the three is `TRUE` ?

(i) Only A (ii) Only B (iii) Only C (iv) Only B and C
(v) None of the above

Solution: (a) (v)
(b) (i)
(c) (iv)
(d) (iii) (Note: if $\neg(p \Rightarrow q)$ is `TRUE` then q is `FALSE` and p is `TRUE` .)
(e) (ii) (Note: it is easy to see that A is not true, e.g., $x = 0.1$; also, C is not true because we can take $x = y = z$.)

Student ID: ____________

Problem 2: [6 pts]

The following three questions are all in the form

"Does a have a multiplicative inverse in Z_b?"

That is, does there exist x in Z_b such that $a \cdot_b x = 1$?

For each question, if the answer is no, prove it.
If the answer is yes, give the value of x (and show your work).

(a) Does 111 have a multiplicative inverse in Z_{1002}?

(b) Does 111 have a multiplicative inverse in Z_{1000}?

(c) Does 111 have a multiplicative inverse in Z_{998}?

Solution:

(a) No. Suppose that such an x existed. Then $111x = 1002q + 1$ for some integer q. Thus $1 = 111x - 1002q = 3(37x - 334q)$ and the right hand side is divisible by 3 while the left hand side isn't.

Alternatively, it is enough to note that $gcd(1002, 111) = 3 > 1$.

(b) Yes. Note that

$$1 \cdot 1000 + (-9) \cdot 111 = 1.$$

Thus, $x = (-9) \bmod 1000 = 991$ is the multiplicative inverse of 111 in Z_{1000}.

(c) Yes. Note that

$$(-1) \cdot 998 + 9 \cdot 111 = 1.$$

Thus, $x = 9$ is the multiplicative inverse of 111 in Z_{998}.

Student ID: ____________

Problem 3: [10 pts]

A die has six sides numbered 1,2,3,4,5,6,

Consider the following game. We roll five (5) dice, each die having a different color; red, blue, green, yellow, orange.
Denote the outcome by an ordered 5-tuple. For example, $(1, 2, 3, 2, 2)$ means that red shows a '1', green shows a '3' and all the rest of the dice show a '2'.

- a *two-of-a-kind* is when two different colored dice show the same number
- a *three-of-a-kind* is when three different colored dice show the same number
- a *four-of-a-kind* is when four different colored dice show the same number
- a *double-two-of-a-kind* is when there are 2 two-of-a-kinds showing different numbers, i.e., not a four-of-a-kind
- a *full-house* is when there is a two-of-a-kind and a three-of-a-kind showing different numbers
- a *straight* is when the five numbers are all different and consecutive

Examples:
$(1, 2, 4, 5, 4)$ contains 1 *two-of-a-kind*, while $(1, 2, 2, 2, 1)$ contains 4 of them.
$(1, 2, 2, 2, 1)$ contains 1 *three-of-a-kind*, while $(3, 3, 1, 3, 3)$ contains 4 of them.
$(2, 3, 5, 2, 5)$ is a *double-two-of-a-kind*, but $(5, 1, 5, 5, 5)$ is not.
$(2, 5, 2, 5, 5)$ is a *full-house*, but $(5, 5, 2, 5, 6)$ is not.
$(3, 1, 5, 4, 2)$ and $(6, 4, 3, 2, 5)$ are *straights*.

Now answer the following questions. It is not necessary to show your work.
The question "*how many ways?*", means "*how many different tuples*".

(a) How many different ways are there of rolling a *full-house*?

(b) How many different ways are there of rolling a *straight*?

(c) How many different ways are there of rolling the dice so you see a *double-two-of-a-kind* but do not see a *three-of-a-kind*?

(d) How many different ways are there of rolling the dice so you see a *two-of-a-kind* but do not see a *three-of-a-kind* and do not see a *double-two-of-a-kind*?

Solution: (a) There are 6 ways of choosing the value of the *two-of-a-kind* and then 5 of choosing the *three-of-a-kind*. After the values are chosen, there are $\binom{5}{3}$ ways of choosing the locations of the triple (and pair). So, the answer is

$$6 \cdot 5 \cdot \binom{5}{3} = 30 \cdot 10 = 300.$$

(b) There are only two types of straights; those that start with a '1' and those that start with a '2'. For each type of straight there are $5! = 120$ different ways of rolling that straight. So the answer is

$$2 \cdot 5! = 240.$$

(c) There are 6 ways of choosing the value of the first pair, then 5 of the next pair and then 4 of the non-pair item. There are $\binom{5}{2}\binom{3}{2}$ ways of choosing the locations of the first pair, second pair and last item. We then have to divide by two because of the symmetry between the first and second pair. This gives

$$6 \cdot 5 \cdot 4 \cdot \binom{5}{2} \cdot \binom{3}{2} \cdot \frac{1}{2} = 1800.$$

(d) In this case, 2 of the dice show the same numbers and the remaining three dice all show different numbers. There are 6 ways of choosing the value of the pair and $\binom{5}{2}$ locations for it. The answer therefore is

$$6 \cdot \binom{5}{2} \cdot 5 \cdot 4 \cdot 3 = 3600.$$

Problem 4: [7 pts]

Prove by induction that

$$\forall n \in Z^+, \qquad 4 \mid (3^{2n-1} + 1),$$

i.e., $(3^{2n-1} + 1)$ is a multiple of 4.

Solution: Base case:

When $n = 1$, $3^{2n-1} + 1 = 3^1 + 1 = 4$. So the base case is true.

Inductive step:

Now consider $n > 1$. We take the inductive hypothesis that the result holds for $n - 1$, i.e., there exists an integer m such that

$$3^{2(n-1)-1} + 1 = 3^{2n-3} + 1 = 4m.$$

We want to prove that the result holds for n.

We rewrite $3^{2n-1} + 1$ as follows:

$$\begin{aligned}
3^{2n-1} + 1 &= 9 \cdot 3^{2n-3} + 1 \\
&= 9 \cdot 3^{2n-3} + 9 - 8 \\
&= 9 \cdot (3^{2n-3} + 1) - 8 \\
&= 9 \cdot 4m - 8 \\
&= 4 \cdot (9m - 2).
\end{aligned}$$

Thus $3^{2n-1} + 1$ is a multiple of 4.

By the weak principle of mathematical induction, we can conclude that $4 \mid (3^{2n-1} + 1)$ for all $n \in Z^+$.

Student ID: ____________

Problem 5: [11 pts]

For this problem you may assume that n is a nonnegative power of 3. Recall that if $f(n)$ and $g(n)$ are functions, to prove that $f(n) = O(g(n))$ you must prove that there exist some $n_0 \geq 0$ and $c > 0$ such that

$$\forall n > n_0, \quad f(n) \leq c g(n)$$

(a) Suppose function $T(n)$ satisfies $T(1) = 9$ and, for $n > 1$

$$T(n) \leq 2T\left(\frac{n}{3}\right) + 4n$$

Prove that $T(n) = O(n)$.

(b) Suppose function $T(n)$ satisfies $T(1) = 9$ and, for $n > 1$

$$T(n) \leq 3T\left(\frac{n}{3}\right) + 4n$$

Prove that $T(n) = O(n \log n)$.

(c) Suppose function $T(n)$ satisfies $T(1) = 9$ and, for $n > 1$

$$T(n) \leq 9T\left(\frac{n}{3}\right) + 4n$$

Prove that $T(n) = O(n^2)$.

Hint: Prove by induction.

Solution: (a) Let $n_0 = 0$. In order for $T(1) \leq cn$ we must have $c \geq 9$.

Now suppose that the statement $T(n) \leq cn$ is correct for all $n = 3^j$, $j = 0, 1, 2, \ldots, i-1$. When $n = 3^i$ we use the inductive hypothesis to get

$$\begin{aligned} T(n) &\leq 2T\left(\frac{n}{3}\right) + 4n \\ &\leq 2\left(c\frac{n}{3}\right) + 4n \\ &= \left(\frac{2c}{3} + 4\right) n \end{aligned}$$

As long as $\frac{2c}{3} + 4 \leq c$, i.e., $12 \leq c$, we then have $T(n) \leq cn$.

Working backwards through the 'proof' we see that it is valid with $n_0 = 0$ and any $c \geq \max(9, 12) = 12$.

(b) We want to show that

$$\forall n > n_0, \quad T(n) \le cn \log n.$$

Note that this statement doesn't make sense without specifying the base of the logarithm. As mentioned in class, it doesn't matter which base we choose. To simplify, we will assume that we mean $\log_3 n$.

Note that there is no c for which $9 = T(1) \le cn \log_3 1 = 0$ so we must set $n_0 \ge 1$. Let's set $n_0 = 1$. Then, we must have

$$T(3) \le c \cdot 3 \cdot \log_3 3 = 3\,c.$$

Since we only know that $T(3) \le 3T(1) + 4 \cdot 3 = 39$, we must require that $3\,c \ge 39$ or $c \ge 13$. Now suppose that the statement $T(n) \le cn \log_3 n$ is correct for all $n = 3^j$, $j = 1, 2, \ldots, i-1$. When $n = 3^i$ we use the inductive hypothesis to get

$$\begin{aligned}
T(n) &\le 3T\left(\frac{n}{3}\right) + 4n \\
&\le 3\left(c\frac{n}{3}\log_3 \frac{n}{3}\right) + 4n \\
&= cn\log_3 n - cn + 4n \\
&= cn\log_3 n + (4-c)n
\end{aligned}$$

As long as $c \ge 4$ this gives $T(n) \le cn \log_3 n$.

Working backwards through the 'proof' we see that it is valid with $n_0 = 1$ and any $c \ge \max(13, 4) = 13$.

(c) It turns out that, similar to the problems seen in class, it is very difficult to use induction to prove

$$\forall n > n_0, \quad T(n) \le cn^2.$$

Instead, following the same method used in class, we will prove

$$\forall n > n_0, \quad T(n) \le c_1 n^2 - c_2 n$$

for some $n_0 \ge 0$ and $c_1, c_2 > 0$. This will imply

$$\forall n > n_0, \quad T(n) \le c_1 n^2.$$

We will set $n_0 = 0$. This will imply $9 = T(1) \le c_1 - c_2$. This will be correct as long as

$$c_1 \ge c_2 + 9.$$

Now suppose that the statement $T(n) \le c_1 n^2 - c_2 n$ is correct for all $n = 3^j$, $j = 0, 1, 2, \ldots, i-1$. When $n = 3^i$ we use the inductive hypothesis to get

$$\begin{aligned} T(n) &\le 9T\left(\frac{n}{3}\right) + 4n \\ &\le 9\left(c_1 \frac{n^2}{9} - c_2 \frac{n}{3}\right) + 4n \\ &= c_1 n^2 - n\left(3c_2 - 4\right) \end{aligned}$$

If $3c_2 - 4 \ge c_2$, i.e., $c_2 \ge 2$, this gives $T(n) \le c_1 n^2 - c_2 n$.
Working backwards through the 'proof' we see that it is valid, e.g., with $n_0 = 0$, $c_2 = 2$ and $c_1 = c_2 + 9 = 11$.

Student ID: ____________

Problem 6: [10 pts]

Let n be a non-negative power of 4 and a a nonnegative integer.
Let $T(n)$ be the function defined by

$$T(n) = \begin{cases} 1 & \text{if } n = 1 \\ aT\left(\frac{n}{4}\right) + n^2 & \text{if } n > 1 \end{cases}$$

(a) Find a closed formula for $T(n)$.
For this part, you should assume that a is a fixed constant.
Your formula should *not* use the summation ($\sum$) symbol.

(b) For which non-negative integer values of a is $T(n) = O(n^2)$?
For which non-negative integer values of a is $T(n) = O(n^2 \log n)$?
For which non-negative integer values of a is $T(n) = O(n^3)$?

For this problem, it is not necessary for you to prove your answer.

Solution: (a) By iterating the recurrence we get

$$\begin{aligned}
T(n) &= aT\left(\frac{n}{4}\right) + n^2 \\
&= a\left(aT\left(\frac{n}{4^2}\right) + \left(\frac{n}{4}\right)^2\right) + n^2 \\
&= a^2T\left(\frac{n}{4^2}\right) + a\left(\frac{n}{4}\right)^2 + n^2 \\
&= a^3T\left(\frac{n}{4^3}\right) + a^2\left(\frac{n}{4^2}\right)^2 + a\left(\frac{n}{4}\right)^2 + n^2 \\
&= \cdots \\
&= a^jT\left(\frac{n}{4^j}\right) + n^2\sum_{i=0}^{j-1}\left(\frac{a}{4^2}\right)^i \\
&= a^{\log_4 n}T(1) + n^2\sum_{i=0}^{\log_4 n-1}\left(\frac{a}{4^2}\right)^i \\
&= n^{\log_4 a} + n^2\sum_{i=0}^{\log_4 n-1}\left(\frac{a}{4^2}\right)^i
\end{aligned}$$

where we used the fact that $a^{\log_4 n} = a^{(\log_4 a)(\log_a n)} = n^{\log_4 a}$.
There are now three cases (the first and third can be combined):
(i) $\frac{a}{4^2} < 1$. In this case $\log_4 a < 2$ and

$$T(n) = n^{\log_4 a} + n^2\frac{1-(a/4^2)^{\log_4 n}}{1-(a/4^2)}.$$

(ii) $\frac{a}{4^2} = 1$, i.e, $a = 16$. In this case $\log_4 a = 2$ and

$$T(n) = n^2 + n^2 \log_4 n.$$

(iii) $\frac{a}{4^2} > 1$. In this case $\log_4 a > 2$ and

$$T(n) = n^{\log_4 a} + n^2 \frac{(a/4^2)^{\log_4 n} - 1}{(a/4^2) - 1}.$$

For the next part we point out that since

$$(a/4^2)^{\log_4 n} = \frac{a^{\log_4 n}}{n^2} = \frac{n^{\log_4 a}}{n^2}$$

we get that both parts are $\Theta(n^{\log_4 a})$.

(b) From the first part we immediately see that
$T(n) = O(n^2)$ if $a < 16$.
$T(n) = O(n^2 \log n)$ if $a \leq 16$.
$T(n) = O(n^3)$ if $\log_4 a \leq 3$, i.e., if $a \leq 64$.

Problem 7: [10 pts]

Suppose we are hashing items into $k \geq 2$ locations, one at a time. Recall that item i causes a *collision* if it is hashed into a location which one of the previous $i-1$ items has already been hashed.

(a) First, suppose we hash n items, with $n \leq k$. What is the probability that all n items hash to different locations?

(b) Now, suppose that we hash items one at a time until the first collision occurs. What is the probability that the ith item causes the first collision?

(c) What is the expected number of items hashed until the first collision occurs? You may express your answer using the summation ($\sum$) sign.

Solution: (a) We denote each outcome of the independent trials process with n trials by an n-tuple, $(x_1, x_2, \ldots, x_n)$, where each $x_i \in \{1, 2, \ldots, k\}$ denotes the location of the ith item.

The size of the sample space of the independent trials process is k^n.

The number of ways in which all n items hash to different locations is $k^{\underline{n}}$.

Thus the probability that all n items hash to different locations is

$$\frac{k^{\underline{n}}}{k^n}.$$

(b) Suppose the ith item in the independent trials process with i trials leads to the first collision. We use an i-tuple, $(x_1, x_2, \ldots, x_i)$, to denote each outcome with each x_j, $1 \leq j \leq i$, defined as in part (a).

The size of the sample space of the independent trials process is k^i.

The number of ways in which the first $(i-1)$ items hash to different locations is $k^{\underline{i-1}}$. After hashing the first $(i-1)$ items to different locations, there are $(i-1)$ ways to hash the ith item, one to each of the $(i-1)$ previously occupied locations. Thus the number of ways in which the ith item leads to the first collision is $(i-1)\, k^{\underline{i-1}}$.

Thus the probability that the ith item is the first collision is

$$\frac{(i-1)\, k^{\underline{i-1}}}{k^i}.$$

(c) Let X denote the random variable for the number of items we hash until the first collision and $P(X = i)$ denote the probability computed in part (b).

The expected number of items can be calculated as

$$\begin{aligned}
E(X) &= \sum_{i=1}^{n} i\, P(X = i) \\
&= \sum_{i=1}^{n} i\, \frac{(i-1)\, k^{i-1}}{k^i} \\
&= \sum_{i=1}^{n} i\, (i-1)\, \frac{k^{i-1}}{k^i} \\
&= \sum_{i=2}^{n} i\, (i-1)\, \frac{k^{i-1}}{k^i}.
\end{aligned}$$

Problem 8: [9 pts]

A test contains 100 true/false questions. Suppose the probability that a student gets the correct answer for each question is 80%, independent of how he did on any other question.

(a) The final score that the student will get on the exam is exactly the number of correct answers he gets (with no penalty for incorrect ones). What are the expected value and variance of his final score?

(b) We now use a different grading scheme which penalizes making an incorrect answer. The final score is calculated by subtracting the number of incorrect answers from the number of correct answers. What are the expected value and variance of the final score based on this grading scheme?

Solution: (a) Let X_i, $1 \leq i \leq 100$, be an indicator random variable with value 0 or 1 to represent the number of correct answers obtained for question i. The expected value and variance of X_i can be calculated as

$$\begin{aligned} E(X_i) &= 1 \cdot 0.8 + 0 \cdot 0.2 \\ &= 0.8 \\ V(X_i) &= 0.8 \cdot (1 - 0.8) = 0.16. \end{aligned}$$

Let X be the random variable for the total number of correct answers. So $X = \sum_{i=1}^{100} X_i$. By the linearity of expectation, we get

$$E(X) = 100 \cdot E(X_i) = 80.$$

Moreover, since X_i and X_j are independent random variables for all $1 \leq i, j, \leq 100$, we have

$$V(X) = 100 \cdot V(X_i) = 16.$$

(b) As in part (a), let X be the random variable for the total number of correct answers. Also, we let Y be the random variable for the total score. So

$$Y = X - (100 - X) = 2X - 100.$$

By the linearity of expectation, we have

$$\begin{aligned} E(Y) &= E(2X - 100) \\ &= E(2X + (-100)) \\ &= E(2X) + E(-100) \\ &= 2 \cdot E(X) - 100 \\ &= 2 \cdot 80 - 100 \\ &= 60. \end{aligned}$$

To compute the variance, we first compute $V(Z+c)$ and $V(cZ)$ for some constant $c \in R$.

$$\begin{aligned}
V(Z+c) &= E[((Z+c)-E(Z+c))^2] \\
&= E[(Z+c-E(Z)-c)^2] \\
&= E[(Z-E(Z))^2] \\
&= V(Z) \\
V(cZ) &= E[(cZ-E(cZ))^2] \\
&= E[(cZ-c\cdot E(Z))^2] \\
&= E[c^2(Z-E(Z))^2] \\
&= c^2 \cdot E[(Z-E(Z))^2] \\
&= c^2 \cdot V(Z).
\end{aligned}$$

Thus we can compute $V(Y)$ as

$$\begin{aligned}
V(Y) &= V(2X-100) \\
&= V(2X) \\
&= 4 \cdot V(X) \\
&= 4 \cdot 16 \\
&= 64.
\end{aligned}$$

Student ID: ____________

Problem 9: [11 pts]

There are n different types of coupon where n is an even number.

Every time you buy something from a department store you will be given a coupon, with equal probability of getting each type of coupon.

Suppose you can get a prize once you have collected $n/2$ coupons of different types.

What is the expected number of times that you need to buy from the department store in order to get a prize?

Your answer should not use the summation ($\sum$) sign, but you can use the notation H_m to denote the m^{th} harmonic number $H_m = \sum_{i=1}^{m}(1/i)$.

Solution: Suppose you have already got $i-1$ types of coupon where $i \geq 1$. The probability that the next coupon is one of a new type is

$$p_i = \frac{n-(i-1)}{n}.$$

Let X_i be the additional number of times to buy from the department store before getting the ith new type of coupon. The expected value of X_i can be calculated as

$$E(X_i) = \frac{1}{p_i} = \frac{n}{n-(i-1)}.$$

Thus the expected number of times to buy from the store before getting a prize is

$$\begin{aligned}
E(X) &= \sum_{i=1}^{n/2} \frac{n}{n-(i-1)} \\
&= n \cdot \sum_{i=1}^{n/2} \frac{1}{n-(i-1)} \\
&= n \cdot \sum_{n/2+1}^{n} \frac{1}{i} \\
&= n \cdot \left(\sum_{i=1}^{n} \frac{1}{i} - \sum_{i=1}^{n/2} \frac{1}{i} \right) \\
&= n \cdot (H_n - H_{n/2}).
\end{aligned}$$

Problem 10: [11 pts]

Five married couples (i.e., 10 people) sit down at random in a row of 10 seats. That is, each one of the 10! different ways of seating the people is equally likely to occur.

We say that a couple *sits together* if the husband and wife in that couple sit next to each other.

(a) What is the probability that $k\,(1 \le k \le 5)$ specified couples end up sitting together (regardless of whether the other $5-k$ couples sit together or not)?

The intent of this question is that if the specified couples are $i_1, i_2, \ldots, i_k$, then couple i_1 is sitting together, couple i_2 is sitting together, ..., and couple i_k is sitting together.

(b) What is the probability that no couple sits together?

You may use the summation ($\sum$) sign and $\binom{n}{m}$ to express your answer.

Solution: (a) There are totally 10! ways to seat the 5 couples.

If a couple sits together, we treat it as one single unit. Thus, for the k specified couples to sit together, we can randomly permute the $(10-k)$ units in $(10-k)!$ different ways. For each permutation, there are 2^k ways to seat the k bound couples. Therefore, the probability is

$$\frac{(10-k)!\,2^k}{10!}.$$

(b) Let E_i denote the event that the ith couple sits together. The probability that at least one couple sits together can be computed using the inclusion-exclusion principle as

$$P\Big(\bigcup_{i=1}^{5} E_i\Big) = \sum_{k=1}^{5}(-1)^{k+1} \sum_{\substack{i_1,i_2,\ldots,i_k:\\ 1\le i_1<i_2<\ldots<i_k\le 5}} P(E_{i_1}\cap E_{i_2}\cap\ldots\cap E_{i_k}).$$

We then apply the result in part (a) to get

$$P\Big(\bigcup_{i=1}^{5} E_i\Big) = \sum_{k=1}^{5}(-1)^{k+1}\binom{5}{k}\frac{(10-k)!\,2^k}{10!}.$$

Thus the probability that no couple sits together is

$$1-P\Big(\bigcup_{i=1}^{5} E_i\Big) = 1-\sum_{k=1}^{5}(-1)^{k+1}\binom{5}{k}\frac{(10-k)!\,2^k}{10!} = \sum_{k=0}^{5}(-1)^{k}\binom{5}{k}\frac{(10-k)!\,2^k}{10!}.$$

Student ID: ____________

Problem 11: [10 pts]

Suppose that we have 100 balls labelled 1-100.
We now play a game and choose 10 of them *without replacement.*
A *good consecutive pair* $(i, i+1)$ is a consecutive pair of labels, both of which have been chosen.
A *good consecutive triple* $(i, i+1, i+2)$ is a consecutive triple of labels, all three of which have been chosen.

The score of the game will be

2 times the number of *good consecutive pairs*
plus
3 times the number of *good consecutive triples*

For example, the set of labels

$$(10, 11, 12, 13, 22, 23, 24, 33, 34, 78)$$

contain the six good consecutive pairs

$$(10, 11),\ (11, 12), (12, 13),\ (22, 23),\ (23, 24),\ (33, 34)$$

and three good consecutive triples

$$(10, 11, 12),\ (11, 12, 13),\ (22, 23, 24),$$

so its score is $2 \times 6 + 3 \times 3 = 21$.

The question you need to answer is:
What is the expected score when 10 balls are chosen randomly, without replacement, from the 100 balls?

For this problem, you should show how you derived your solution. You should give your solution as a closed formula, but it is not necessary to write down an actual number for the solution.

Solution: Let I_i, $i = 1, \ldots, 99$, be the indicator random variable for the event that the pair $(i, i+1)$ is chosen. Then

$$E(I_i) = Pr((i, i+1) \text{ is chosen}) = \frac{\binom{98}{8}}{\binom{100}{10}} = \frac{10! \cdot 98!}{8! \cdot 100!} = \frac{10 \cdot 9}{100 \cdot 99} = \frac{1}{10 \cdot 11}$$

Now let J_i, $i = 1, \ldots, 98$, be the indicator random variable for the event that the triple $(i, i+1, i+2)$ is chosen. Then

$$E(J_i) = Pr((i, i+1, i+2) \text{ is chosen}) = \frac{\binom{97}{7}}{\binom{100}{10}} = \frac{10! \cdot 97!}{7! \cdot 100!} = \frac{10 \cdot 9 \cdot 8}{100 \cdot 99 \cdot 98}$$

Let X denote the score. Then

$$X = 2\sum_{i=1}^{99} I_i + 3\sum_{i=1}^{98} J_i.$$

So, by the linearity of expectation

$$\begin{aligned}
E(X) &= E\left(2\sum_{i=1}^{99} I_i + 3\sum_{i=1}^{98} J_i\right) \\
&= 2\cdot 99\cdot E(I_1) + 3\cdot 98\cdot E(J_i) \\
&= 2\cdot 99\frac{\binom{98}{8}}{\binom{100}{10}} + 3\cdot 98\frac{\binom{97}{7}}{\binom{100}{10}}
\end{aligned}$$